Appendix

Appendix A1.1 Study characteristics: O'Connor, 1999, Study A: Intensive Professional Development (quasi-experimental design)

Characteristic	Description
Study citation	O'Connor, R. E. (1999). Teachers Learning Ladders to Literacy. *Learning Disabilities Research & Practice, 14*(4), 203–214. (Study A: Intensive Professional Development)
Participants	Two schools volunteered to implement the intervention and two schools were recruited with a financial incentive (contributions to their school libraries) to serve as controls. Ten classes of 8 to 22 students participated in the study. Six of the ten teachers taught in general Kindergarten classes that included at least three children with disabilities. Two teachers taught in transition Kindergartens composed of students repeating Kindergarten (with or without diagnosed disabilities). Two teachers taught in special education classrooms. All classes included at least two children with disabilities. The children were predominantly European-American and African-American. Students were divided by ability level and labeled as typical learners or at-risk learners.[1] Pretest equivalence was not established for at-risk learners so this subgroup (including the two special education classes) was excluded from the review.[2] Therefore, this intervention report focuses on the findings reported only for the typical learners. The analysis sample of typical learners included 64 students in the intervention group and 41 students in the comparison group across general education and transition classes.
Setting	The study took place in four schools in an urban district.
Intervention	In addition to their typical pre-reading instruction, children in intervention classes were given more than twenty activities from the *Ladders to Literacy* book, including sound isolation, first sounds, rhyming pictures, rhyming, onsets and rhymes with first letters, invented spelling, story grammar, and integrating spelling and reading. The district-sponsored pre-reading curriculum included reading and discussing Big Books, learning letters of the alphabet and common sounds, and practicing writing of letters.
Comparison	Children in comparison classes received the same district-sponsored pre-reading curriculum as intervention classes. Children were also introduced to the concept of rhyme. Activities requiring blending or segmenting (beyond the initial sound) were not observed in any of the classes. The students were matched to the intervention students on the Peabody Picture Vocabulary Test and demographic variables.
Primary outcomes and measurement	For both pretest and posttest, the author administered the Peabody Picture Vocabulary test, the test of Short Term Memory, the Woodcock-Johnson Test of Achievement: Letter-Word Identification subtest, and four tasks: Rhyme Production, Segmenting, Blending, and Rapid Letter Naming. The Dictation subtest of the Woodcock Johnson Tests of Achievement was also used in the study but is not included because it is outside the scope of this Beginning Reading review. (See Appendices A2.1–2.3 for more detailed descriptions of outcome measures.)
Teacher training	In this intensive model of professional development, 14 days of teacher training were spread across the school year. Teachers discussed implementation of program activities, solved issues with materials, and shared data on the progress of their students. Teachers modeled instruction and rehearsed upcoming activities. Researchers worked with teachers to determine appropriate timing of activities and often observed students directly to supply the rationale for the next set of activities.

1. At-risk learners were defined as children with low skills (children with high-incidence disabilities or whose standard scores fell below 85 at PPVT pretest).
2. The groups are deemed non-equivalent if the pretest standardized mean difference is 0.5 or larger. See the Beginning Reading Protocol for more information.

Appendix A1.2 Study characteristics: O'Connor, 1999, Study B: Traditional Professional Development (quasi-experimental design)

Characteristic	Description
Study citation	O'Connor, R. E. (1999). Teachers Learning Ladders to Literacy. *Learning Disabilities Research & Practice, 14*(4), 203–214. (Study B: Traditional Professional Development)
Participants	All the Kindergarten teachers in a district participated and were assigned to an in-service or comparison condition by geographical location, with all eight comparison classes coming from the same school. Students in each classroom were pre-tested and divided by ability level into typical learners and at-risk learners. In the combined analysis sample that included both subgroups, 192 students from nine classrooms were in the treatment group and 126 students from eight classrooms were in the comparison group.[1] The children were predominantly of European-American descent. Class sizes ranged from 18–28 students and all included at least one child with a disability.
Setting	The study took place in a large rural midwestern district.
Intervention	In addition to their typical pre-reading instruction, children in nine intervention classes were given more than twenty *Ladders to Literacy* activities, including sound isolation, first sounds, rhyming pictures, rhyming, onsets and rhymes with first letters, invented spelling, story grammar, and integrating spelling and reading. The district-sponsored pre-reading curriculum included reading and discussing Big Books, learning letters of the alphabet and common sounds, and practicing writing of letters.
Comparison	Children in eight comparison classes received the same district-sponsored pre-reading curriculum as the intervention group and were introduced to the concept of rhyme. The eight comparison classes formed a Kindergarten center in one school. The teachers of these classes routinely planned their instruction together and shared materials. The students were matched on the Peabody Picture vocabulary Test and demographic variables.
Primary outcomes and measurement	For both pretest and posttest, the author administered the Peabody Picture Vocabulary test, the test of Short Term Memory, the Letter-Word Identification subtest of the Woodcock-Johnson Test of Achievement, and the Rhyme Production, Segmenting, Blending, and Rapid Letter Naming tests. The Dictation subtest of the Woodcock Johnson Tests of Achievement was also used in the study but is not included because it is outside the scope of this Beginning Reading review. (See Appendices A2.1–2.3 for more detailed descriptions of outcome measures.)
Teacher training	Teacher participants included general education Kindergarten teachers, Title I teachers and other supporting staff. In this traditional model of professional development, teacher training totaled three and a half days spaced across the school year. Teachers discussed implementation of activities, solved issues with materials, and shared data on the progress of their students. Training provided teachers with modeling and rehearsal of upcoming activities.

1. Findings separated for at-risk learners and typical learners are presented in Appendices A4.1–4.2

Appendix A1.3 Study characteristics: O'Connor, Notari-Syverson, & Vadasy, 1996 (quasi-experimental design)

Characteristic	Description
Study citation	O'Connor, R., Notari-Syverson, A., & Vadasy, P. F. (1996). *The effect of kindergarten phonological intervention on the first grade reading and writing of children with mild disabilities.* Paper presented at the meeting of the American Educational Research Association, New York, NY. (ERIC Document Reproduction Service No. ED394129)
Participants	Three general Kindergarten teachers and two transition teachers (of children repeating Kindergarten) agreed to participate.[1] The two transition teachers were randomly assigned to treatment or control conditions. The three regular classroom teachers were not randomly assigned to the treatment or comparison condition. Students were matched by type of classroom (general or repeating Kindergarteners). In the combined analysis 42 students from three classrooms were in the intervention group and 24 students from two classrooms were in the comparison group.[2] Intervention and comparison students had comparable performance on pretests. The ethnic distribution in the school district, reflected in the composition of three general classes, was 52% African-American, 46% Caucasian, and 2% other ethnicities. The distribution in the transition classrooms was 65% African-American, 25% Caucasian, and 10% other ethnicities. Transition classes for children who repeat Kindergarten had reduced class sizes (12 to 15 students compared to 21 to 25 students in the regular Kindergartens). Longitudinal findings at the end of first grade are presented in Appendix A4.3.
Setting	The study took place in a large, urban school district.
Intervention	The intervention was a supplement to a normal pre-reading instruction. Children in three intervention classes were given twenty-five *Ladders to Literacy* activities over the 6-month intervention period. In the first two months activities stimulated word and syllable awareness. The third and fourth months focused on rhyming, first sound isolation, and onset-rime level blending and segmenting. Letters and sounds were added to phonological activities in the final two months, when children were shown how to use a letter sound to match pictures that start the same. At this point the auditory blending games became more sophisticated, separating each spoken phoneme. In the two general Kindergartens, teachers conducted these activities in short sessions (5 to 15 minutes long) with their whole group of 21 to 25 students. In the transition class, the teachers and assistants usually conducted activities in smaller groups of 3 to 6 students.
Comparison	Children in two comparison classes received the same district-wide pre-reading curriculum as the intervention group. They did not practice auditory blending, segmenting words beyond the first letter, or selecting letters to represent sounds. The comparison transition class included finger-point reading of Big Books in addition to the general curriculum.
Primary outcomes and measurement	For both pretest and posttest, the authors administered the Peabody Picture Vocabulary test, the Letter-Word Identification subtest of the Woodcock-Johnson Test of Achievement, and also Rhyme Production, Segmenting, Blending, First Sound, Sound Repetition, and Rapid Letter Naming tests. For the follow-up tests conducted at the end of first grade, the authors administered a segmentation measure and the Word Attack and Word Identification subtests of the Woodcock-Johnson Test of Achievement. The Dictation subtest of the Woodcock Johnson Tests of Achievement and Test of Written Spelling were also used in the study but are not included because they are outside the scope of this Beginning Reading review. (See Appendices A2.1–2.3 for more detailed descriptions of outcome measures.)
Teacher training	Teachers in the intervention condition received 10 in-service training sessions spaced over the school year. Sessions reviewed materials, discussed the conceptual basis for each activity, and offered practical suggestions for incorporating activities into the class routines and feedback on implementation of earlier activities. Bi-weekly visits from the study authors extended teacher training by addressing particular classroom concerns, modifying activities, and monitoring program implementation.

1. Two self-contained classes of Kindergarteners with mild disabilities also participated in the study. Because appropriate controls did not exist for children in the self-contained classes, these classes were not in included in the review.
2. The analyses excluded children with disabilities in the general classrooms.

Appendix A1.4 Study characteristics: Fuchs et al., 2001 (randomized control trial with randomization problems)

Characteristic	Description
Study citation	Fuchs, G., Fuchs, L. S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N. J., Braun, M., and O'Conner, R. E. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology 93*(2), 251–267.
Participants	Thirty-three teachers were stratified and then randomly assigned to three conditions: *Ladders to Literacy, Ladders to Literacy plus Peer-Assisted Learning Strategies (PALS),* and a comparison condition. Students were selected to participate based on the Rapid Letter Naming test and student names were presented to teachers for their review and adjustment.[1] For rating purposes, the WWC focused on the 11 teachers with 136 students that were in the *Ladders to Literacy* treatment group and 11 teachers with 135 students that were in the comparison group.[2] Students in both conditions were also compared longitudinally and in terms of varying levels of reading performance (high, medium, and low)[3] in fall of first grade.[4]
Setting	The study took place in four Title I and four non-Title I schools in the Nashville public school system.
Intervention	Children in the intervention classes received their typical pre-reading instruction and were given 15 *Ladders to Literacy* activities for a maximum of 45 minutes a week for twenty weeks. These activities included word and syllable awareness, rhyming, first sound isolation, onset-rhyme blending, sound segmentation, journal writing, "letter sound of the week," "morning message," nursery rhymes and poems, and shared storybook reading. Only three of the activities presented students with printed letters. Students in the *Ladders to Literacy plus PALS* classes participated in a 20-week phonological awareness training and beginning decoding curriculum. The *PALS* component, which was implemented for 16 weeks, required children to work in pairs with peers of their own ages. The *PALS* activities focused on the correct sounds of letters and required children to read aloud sight words, decodable words, and simple sentences.
Comparison	Students in the comparison classes received their regular whole-class reading instruction. Nearly two-thirds of teachers used the school district's formally adopted text: the Harcourt-Brace *Treasury of Literature: First Street Collection for Kindergarten.* A majority of the teachers used *First Street*'s Big Books, and about half of the teachers reported using *High Hat.* A majority of comparison teachers taught alphabet letter naming.
Primary outcomes and measurement	For both pretest and posttest, the authors administered two subtests of the Woodcock Reading Mastery test (Word Identification and Word Attack), a rapid letter sound test, and a segmentation task. At posttest, the authors also administered a blending task. All these tests were also used during follow-up testing which occurred in the fall of first grade. The Spelling subtest of the Wechsler Individual Achievement test was also used in the study but is not included because it is outside the scope of this Beginning Reading review. (See Appendix A2.1–2.3 for more detailed descriptions of outcome measures.)
Teacher training	Intervention teachers attended a full-day workshop that included discussion of phonological awareness tasks and description of the 15 *Ladders to Literacy* activities that teachers were asked to implement. *Ladders to Literacy* and *PALS* teachers attended an additional half-day workshop to prepare their students in *PALS.*

1. Although teachers were randomly assigned to a treatment or intervention condition, teachers' judgment played a role in selecting students for the analysis sample. The WWC could not verify that student selection for the analysis sample was unrelated to treatment status, so the study met WWC evidence standards with reservations.
2. Because it involved a variation of the intervention, the *Ladders to Literacy plus PALS* (phonological awareness plus beginning decoding) portion of the study is not included in this report.
3. Reading performance was designated by the results of the Rapid Letter Naming pretest and teacher review/adjustment.
4. These results are presented in Appendices A4.

Appendix A2.1 Outcome measures in the alphabetics domain by construct

Outcome measure	Description
Phonemic Awareness	
Test of Short Term Memory	This test measured children's ability to reproduce phonemes that varied from two to four phonemes in length (as cited in O'Connor, 1999).
Phonological Awareness	
Segmentation/segmenting	Segmentation with a test, similar to the Yopp-Singer test, that asks students to deconstruct words into component sounds (e.g. to say the sounds in cat) (as cited in Fuchs et al., 2001, and O'Connor et al., 1996).
Blending	Blending was tested using 22 three-sound words (e.g., soap, food, mom). The number of words correctly blended from sounds into a word in one minute was recorded (as cited in Fuchs et al., 2001). Correct responses were awarded two points and the examiner provided the answer for missed words. Next, the examiner presented any missed items in onset-rime format (s—oap), and correct responses were awarded one point (as cited in O'Connor et al., 1996).
Rhyme production	This test measured students' ability to rhyme one-syllable words (as cited in O'Connor, 1999; O'Connor et al., 1996).
First sound	Items were scored correct if the child provided only the first sound of the word (e.g., for "pill", /p/ or /puh/ was correct; /pi/ was not) (as cited in O'Connor et al., 1996).
Sound repetition	For this test, children repeated isolated phonemes separated with a half-second pause ("Say exactly what I say :/p/ /I/ /f/"). Items varied from two to four in length, with the number of correct items reported (as cited in O'Connor, Notari-Syverson, and Vadasy, 1996).
Letter Knowledge	
Rapid letter naming	This test determines the number of letters a child can name correctly in one minute (as cited in Fuchs et al., 2001; O'Connor et al., 1996).
Phonics	
Woodcock Reading Mastery Test: Word Attack subtest	The standardized test measures phonemic decoding skills by asking students to read pseudowords. Students are aware that the words are not real (as cited in Fuchs et al., 2001; O'Connor et al., 1996).
Woodcock Johnson Tests of Achievement: Letter-Word Identification subtest.	In this subtest, the examiner showed children letters or words, and children said the letter name or read the word (as cited in O'Connor, 1999; O'Connor et al., 1996).
Rapid letter sound	The test assesses the number of letter-sounds a student can name correctly in one minute. Students are shown a sheet with lower case letters and asked to tell the sound each letter makes. It was adapted from the work of Levy and Lysynchuk (as cited in Fuchs et al., 2001).

Appendix A2.2 Outcome measures in the fluency domain

Outcome measure	Description
Test of Oral Reading Fluency	The test measures words read correctly in one minute on a primer level passage (as cited in O'Connor et al., 1996).

Appendix A2.3 Outcome measures in the comprehension domain

Outcome measure	Description
Peabody Picture Vocabulary Test–R	This individually administered norm-referenced test measured receptive vocabulary. Children are shown four pictures and asked to point to the one described by the examiner (as cited in O'Connor, 1999; O'Connor et al., 1996).

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain by construct[1]

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Construct: Phonemic awareness								
O'Connor, 1999; Study A: Intensive Professional Development (quasi-experimental design)[8]								
Short term memory	Kindergarten (typical learners)	8/105	11.69 (1.34)	10.23 (1.51)	1.46	1.03	Statistically significant	+35
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
Short term memory	Kindergarten	17/ 318	9.95 (1.69)	9.61 (1.93)	0.34	0.19	ns	+8
Construct: Phonological awareness								
O'Connor, 1999; Study A: Intensive Professional Development (quasi-experimental design)[8]								
Rhyme production	Kindergarten (typical learners)	8/105	9.87 (1.32)	9.44 (2.01)	0.43	0.26	ns	+10
Segmentation	Kindergarten (typical learners)	8/105	22.71 (6.72)	10.69 (6.71)	12.02	1.78	Statistically significant	+46
Blending	Kindergarten (typical learners)	8/105	16.24 (4.55)	11.89 (6.14)	4.35	0.83	Statistically significant	+30
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
Rhyme production	Kindergarten	17/318	5.18 (4.06)	4.22 (4.13)	0.96	0.23	ns	+9
Segmentation	Kindergarten	17/318	20.63 (9.64)	11.39 (10.22)	9.24	0.93	Statistically significant	+32
Blending	Kindergarten	17/318	13.07 (5.38)	10.18 (5.78)	2.89	0.52	ns	+20
O'Connor et al, 1996 (quasi-experimental design)[8]								
Sound repetition	Kindergarten	5/66	10.90 (1.20)	10.00 (1.60)	0.90	0.66	ns	+24

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain by construct[1] *(continued)*

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Blending	Kindergarten	5/66	15.50 (5.30)	11.20 (6.9)	4.30	0.72	ns	+26
First sound	Kindergarten	5/66	10.30 (1.50)	8.90 (2.80)	1.40	0.67	ns	+25
Segmenting	Kindergarten	5/66	23.70 (7.90)	9.50 (5.8)	14.20	1.94	Statistically significant	+47
Rhyme production	Kindergarten	5/66	9.60 (1.6)	9.10 (2.40)	0.50	0.26	ns	+10
Construct: Letter knowledge								
O'Connor, 1999; Study A: Intensive Professional Development (quasi-experimental design)								
Rapid letter naming	Kindergarten (typical learners)	8/105	45.50 (11.77)	36.32 (14.95)	9.18	0.70	ns	+26
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)								
Rapid letter naming	Kindergarten	17/318	36.67 (17.19)	31.51 (18.42)	5.16	0.29	ns	+11
O'Connor et al, 1996 (quasi-experimental design)								
Rapid Letter naming	Kindergarten	5/66	38.80 (13.10)	35.70 (15.40)	3.10	0.22	ns	+9
Construct: Phonics								
O'Connor, 1999; Study A: Intensive Professional Development (quasi-experimental design)								
WJ Letter-Word Identification	Kindergarten (typical learners)	8/105	16.14 (3.44)	13.00 (1.52)	3.14	1.09	Statistically significant	+36
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)								
WJ Letter-Word Identification	Kindergarten	17/318	13.45 (2.96)	12.08 (3.55)	1.37	0.43	ns	+17

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain by construct[1] *(continued)*

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
O'Connor et al, 1996 (quasi-experimental design)								
WJ Letter-Word Identification	Kindergarten	5/66	99.30 (11.20)	96.10 (10.30)	3.20	0.29	ns	+11
Fuchs et al., 2001[9] (randomized controlled trial with randomization problems)								
Rapid letter sound	Kindergarten	22/271	16.99 (nr)	15.81 (nr)	1.18	na[10]	ns	na[10]
WJ Word Attack	Kindergarten	22/271	3.32 (nr)	2.03 (nr)	1.29	na[11]	ns	na[11]
WJ Word Identification	Kindergarten	22/271	7.12 (nr)	5.47 (nr)	1.65	na[12]	ns	na[12]
Average[13] for alphabetics domain (O'Connor, 1999, Study A)						0.95	Statistically significant	+33
Average[13] for alphabetics domain (O'Connor, 1999; Study B)						0.43	Statistically significant	+17
Average[13] for alphabetics domain (O'Connor et al., 1996)						0.68	ns	+25
Average[13] for alphabetics domain (Fuchs et al., 2001)						na	ns	na
Domain average[13] for alphabetics						0.69	na	+25

ns = not statistically significant
nr = not reported
na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement index. Subgroup findings and findings from the alternative intervention group from the same studies are not included in these ratings, but are reported in Appendices A4.1–A4.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Ladders to Literacy* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain by construct[1] *(continued)*

8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formula the WWC used to calculate statistical significance.
    In the case of O' Connor (1999) and O'Connor (1996), corrections for clustering and multiple comparisons were needed so the significance levels may differ from those reported in the original study.
    In the case of Fuchs et al. (2001), a correction for multiple comparisons was needed so the significance levels may differ from those reported in the original study.
9. Student-level data were not available to the WWC. The authors' analysis provides a conservative test of the intervention's effectiveness. The study contributed to the intervention ratings through the statistical significance of findings.
10. Student-level standard deviations were not available for this study. Cluster-level standard deviations were 8.39 for the intervention group and 6.20 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations.
11. Student-level standard deviations were not available for this study. Cluster-level standard deviations were 3.46 for the intervention group and 2.54 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations.
12. Student-level standard deviations were not available for this study. Cluster-level standard deviations were 6.71 for the intervention group and 4.76 for the comparison group. Because the student-level effect size and improvement index could not be computed, the magnitude of the effect size was not considered for rating purposes. However, the statistical significance for this study is comparable to other studies and is included in the intervention rating. For further details, please see Technical Details of WWC-Conducted Computations
13. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the fluency domain[1]

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
O'Connor et al, 1996 (quasi-experimental design)[7]								
Fluency	Kindergarten	5/ 66	66.70 (35.5)	42.50 (29.00)	24.20	0.72	ns	+26
Domain average[8] for fluency						0.72	ns	+26

ns = not statistically significant

1. This appendix reports findings at the end of first grade considered for the effectiveness rating and the improvement index.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formula the WWC used to calculate statistical significance. In the case of O' Connor et al (1996), a correction for clustering was needed so the significance level may differ from those reported in the original study.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places.

Appendix A3.3 Summary of study findings included in the rating for the comprehension domain[1]

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Construct: Vocabulary Development								
O'Connor, 1999; Study A: Intensive Professional Development (quasi-experimental design)[8]								
PPVT	Kindergarten (typical learners)	8/105	111.04 (13.74)	104.79 (13.76)	6.25	0.45	ns	+17
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
PPVT	Kindergarten (typical learners)	17/318	100.60 (15.36)	100.39 (16.15)	0.21	0.01	ns	+1
O'Connor et al., 1996; (quasi-experimental design)[8]								
PPVT	Kindergarten (children without disabilities)	5/66	103.80 (14.10)	100.20 (16.10)	3.60	0.24	ns	+9
Domain average[9] for comprehension						0.23	ns	+9

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Subgroup findings from the same studies are not included in these ratings, but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Ladders to Literacy* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formula the WWC used to calculate statistical significance.
   In the case of the O' Connor (1999) and O'Connor (1996) studies, corrections for clustering were needed, so the significance levels may differ from those reported in the original study.
9. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

Appendix A4.1 Summary of subgroup findings for alphabetics domain by construct[1]

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Construct: Phonemic Awareness								
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
Short term memory	Kindergarten (at-risk learners[9])	17/95	9.15 (2.08)	8.43 (2.36)	0.72	0.33	ns	+13
Short term memory	Kindergarten (typical learners)	17/223	10.29 (1.36)	10.10 (1.48)	0.19	0.13	ns	+5
Construct: Phonological Awareness								
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
Rhyme production	Kindergarten (at-risk learners)	17/95	5.58 (4.40)	3.86 (4.29)	1.72	0.39	ns	+15
Segmentation	Kindergarten (at-risk learners)	17/95	17.59 (11.39)	6.46 (8.61)	11.13	1.06	Statistically significant	+36
Blending	Kindergarten (at-risk learners)	17/95	8.89 (5.16)	7.46 (4.78)	1.43	0.28	ns	+11
Rhyme production	Kindergarten (typical learners)	17/223	6.89 (3.91)	6.26 (4.07)	0.63	0.16	ns	+6
Segmentation	Kindergarten (typical learners)	17/223	21.95 (8.49)	13.44 (10.17)	8.51	0.92	Statistically significant	+32
Blending	Kindergarten (typical learners)	17/223	14.88 (4.39)	11.31 (5.81)	3.57	0.71	Statistically significant	+26
Fuchs et al., 2001 (randomized controlled trial with randomization problems)[8]								
Segmentation	Kindergarten (low achievers)	22/nr[10]	11.96 (nr/6.71)[11]	8.93 (nr/7.67)[10]	3.03	na	ns	na
Segmentation	Kindergarten (average achievers)	22/nr[10]	15.76 (nr/6.73)[10]	7.99 (nr/5.74)[10]	7.77	na	Statistically significant	na

(continued)

Appendix A4.1 Summary of subgroup findings for alphabetics domain by construct *(continued)*

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Segmentation	Kindergarten (high achievers)	22/nr10	22.96 (nr/6.93)[10]	14.16 (nr/8.56)[10]	8.80	na	Statistically significant	na
Blending	Kindergarten (low achievers)	22/nr[10]	6.40 (nr/4.16)[10]	2.48 (nr/2.85)[10]	3.92	na	Statistically significant	na
Blending	Kindergarten (average achievers)	22/nr[10]	7.91 (nr/3.56)[10]	3.52 (nr/3.18)[10]	4.39	na	Statistically significant	na
Blending	Kindergarten (high achievers)	22/nr[10]	12.02 (nr/4.54)[10]	6.50 (nr/5.07)[10]	5.52	na	Statistically significant	na
Construct: Letter Knowledge								
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
Rapid letter naming	Kindergarten (at-risk learners)	17/95	26.32 (15.75)	18.11 (13.46)	8.21	0.55	ns	+21
Construct: Phonics								
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
WJ Letter-Word Identification	Kindergarten (at-risk learners)	17/95	11.94 (3.06)	9.78 (3.18)	2.16	0.69	Statistically significant	+25
WJ Letter-Word Identification	Kindergarten (typical learners)	17/223	14.10 (2.68)	13.04 (3.25)	1.06	0.36	ns	+14
Fuchs et al., 2001 (randomized controlled trial with randomization problems)[8]								
Rapid letter sound	Kindergarten (low achievers)	22/nr[10]	9.89 (nr/6.77)[10]	8.49 (nr/5.72)[10]	1.40	na	ns	na
Rapid letter sound	Kindergarten (average achievers)	22/nr[10]	9.98 (nr/7.46)[10]	14.28 (nr/7.89)[10]	–4.30	na	ns	na
Rapid letter sound	Kindergarten (high achievers)	22/nr[10]	24.72 (nr/9.23)	24.65 (nr/7.43)[10]	0.07	na	ns	na
WJ Word Attack	Kindergarten (low achievers)	22/nr[10]	0.61 (nr/0.98)[10]	0.41 (nr/1.00)[10]	0.20	na	ns	na

(continued)

Appendix A4.1 Summary of subgroup findings for alphabetics domain by construct *(continued)*

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
WJ Word Attack	Kindergarten (average achievers)	22/nr[10]	2.47 (nr/4.30)[10]	1.38 (nr/2.02)[10]	1.09	na	ns	na
WJ Word Attack	Kindergarten (high achievers)	22/nr[10]	6.88 (nr/6.49)[10]	4.29 (nr/5.16)[10]	2.59	na	ns	na
WJ Word Identification	Kindergarten (low achievers)	22/nr[10]	1.34 (nr/1.28)[10]	1.00 (nr/1.34)[10]	0.34	na	ns	na
WJ Word Identification	Kindergarten (average achievers)	22/nr[10]	5.63 (nr/8.47)[10]	2.87 (nr/4.59)[10]	2.76	na	ns	na
WJ Word Identification	Kindergarten (high achievers)	22/nr[10]	14.42 (nr/12.24)[10]	12.54 (nr/10.01)[10]	1.88	na	ns	na

ns = not statistically significant
nr = not reported
na = not applicable

1. This appendix presents subgroup findings—at-risk learners and typical learners (O'Connor, 1999) and low, average, and high achievers (Fuchs et al., 2001)—for measures that fall in the alphabetics domain. Total group scores were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Ladders to Literacy* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance.
   In the case of O'Connor (1999; B), corrections for clustering were needed, so the significance levels may differ from those reported in the original study.
   In the case of Fuchs et al. (2001), no corrections were needed.
9. At-risk learners were defined as children with low skills (i.e., children with high-incidence disabilities or whose standard scores fell below 85 at PPVT pretest).
10. The total number of students was 271 with approximately 40% low achievers, 30% average achievers, and 30 high achievers; however, the exact number in each group was not provided.
11. Student-level standard deviations were not available for this study. Thus, a student-level effect size and improvement index could not be computed. Cluster-level standard deviations are presented in the author's findings columns.

Appendix A4.2 Summary of subgroup findings for comprehension domain[1]

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Construct: Vocabulary Development								
O'Connor, 1999; Study B: Traditional Professional Development (quasi-experimental design)[8]								
PPVT	Kindergarten (at-risk learners)	17/95	93.06 (14.98)	90.66 (16.90)	2.40	0.15	ns	+6
PPVT	Kindergarten (typical learners)	17/223	103.87 (14.39)	104.43 (14.05)	–0.56	–0.04	ns	–2

ns = not statistically significant

1. This appendix presents subgroup findings (at-risk learners and typical learners) for measures that fall in comprehension domain. Total group scores were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Ladders to Literacy* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance.

   In the case of O'Connor (1999; B), corrections for clustering were needed, so the significance levels may differ from those reported in the original study.

Appendix A4.3 Summary of longitudinal findings for the alphabetics domain by construct[1]

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Construct: Phonological Awareness								
O'Connor et al, 1996 (quasi-experimental design)—End of first grade[6]								
Segmentation	Kindergarten	5/66	31.40 (12.40)	27.50 (8.50)	3.90	0.35	ns	+14
Fuchs et al., 2001 (randomized controlled trial with randomization problems)—Fall of first grade[6]								
Segmentation	Kindergarten (low achievers)	22/nr[8]	21.23 (nr/6.12)[7]	16.38 (nr/7.11)[7]	4.85	na	ns	na
Segmentation	Kindergarten (average achievers)	22/nr[8]	23.79 (nr/8.49)[7]	18.42 (nr/6.92)[7]	5.37	na	ns	na
Segmentation	Kindergarten (high achievers)	22/nr[8]	25.64 (nr/8.82)[7]	23.47 (nr/7.70)[7]	2.17	na	ns	na
Segmentation	Kindergarten (across categories)	22/271	23.56 (nr/6.32)[7]	19.42 (nr/4.55)[7]	4.14	na	ns	na
Blending	Kindergarten (low achievers)	22/nr[8]	14.14 (nr/2.66)[7]	9.22 (nr/3.60)[7]	4.92	na	Statistically significant	na
Blending	Kindergarten (average achievers)	22/nr[8]	14.47 (nr/5.06)[7]	12.47 (nr/3.37)[7]	2.00	na	ns	na
Blending	Kindergarten (high achievers)	22/nr[8]	16.78 (nr/4.33)[7]	14.21 (nr/4.95)[7]	2.57	na	ns	na
Blending	Kindergarten (across categories)	22/271	15.13 (nr/3.29)[7]	11.96 (nr/3.18)[7]	3.17	na	Statistically significant	na
Construct: Phonics								
O'Connor et al, 1996 (quasi-experimental design)—End of first grade[6]								
WJ Letter-Word Identification	Kindergarten	5/66	111.40 (14.80)	109.8 (17.00)	1.60	0.10	ns	+4
WJ Word attack	Kindergarten	5/66	118.40 (13.00)	108.9 (14.30)	9.50	0.70	ns	+26

(continued)

Appendix A4.3 Summary of longitudinal findings for the alphabetics domain by construct[1] *(continued)*

Outcome measure	Study sample	Sample size (teachers/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Ladders to Literacy* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Ladders to Literacy* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Fuchs et al., 2001 (randomized controlled trial with randomization problems)—Fall of first grade[6]								
Rapid letter sound	Kindergarten (low achievers)	22/nr[8]	21.25 (nr/6.11)[7]	18.81 (nr/7.26)[7]	2.44	na	ns	na
Rapid letter sound	Kindergarten (average achievers)	22/nr[8]	29.18 (nr/9.87)[7]	22.33 (nr/7.91)[7]	6.85	na	ns	na
Rapid letter sound	Kindergarten (high achievers)	22/nr[8]	30.53 (nr/10.05)[7]	34.24 (nr/7.81)[7]	–3.71	na	ns	na
Rapid letter sound	Kindergarten (across categories)	22/271	26.98 (nr/6.39)[7]	25.13 (nr/6.12)[7]	1.85	na	ns	na
WJ Word Attack	Kindergarten (low achievers)	22/nr[8]	3.46 (nr/2.32)[7]	1.94 (nr/2.14)[7]	1.52	na	ns	na
WJ Word Attack	Kindergarten (average achievers)	22/nr[8]	5.56 (nr/5.13)[7]	3.03 (nr/2.14)[7]	2.53	na	ns	na
WJ Word Attack	Kindergarten (high achievers)	22/nr[8]	8.85 (nr/7.61)[7]	8.47 (nr/9.14)[7]	0.38	na	ns	na
WJ Word Attack	Kindergarten (across categories)	22/271	5.96 (nr/4.70)[7]	4.48 (nr/4.31)[7]	1.48	na	ns	na
WJ Word Identification	Kindergarten (low achievers)	22/nr[8]	10.20 (nr/6.00)[7]	8.21 (nr/5.66)[7]	1.99	na	ns	na
WJ Word Identification	Kindergarten (average achievers)	22/nr[8]	15.96 (nr/8.09)[7]	13.95 (nr/7.45)[7]	2.01	na	ns	na
WJ Word Identification	Kindergarten (high achievers)	22/nr[8]	26.57 (nr/13.33)[7]	28.04 (nr/13.54)[7]	–1.47	na	ns	na
WJ Word Identification	Kindergarten (across categories)	22/271	17.57 (nr/8.61)[7]	16.73 (nr/6.76)[7]	0.84	na	ns	na
PPVT	Kindergarten (at-risk learners)	17/95	93.06 (14.98)	90.66 (16.90)	2.40	0.15	ns	+6

(continued)

Appendix A4.3 Summary of longitudinal findings for the alphabetics domain by construct[1] *(continued)*

			Authors' findings from the study					
			Mean outcome (standard deviation[2])		WWC calculations			
Outcome measure	Study sample	Sample size (teachers/ students)	*Ladders to Literacy* group[3]	Comparison group	Mean difference[4] (*Ladders to Literacy* – comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
PPVT	Kindergarten (typical learners)	17/223	103.87 (14.39)	104.43 (14.05)	–0.56	–0.04	ns	–2

ns = not statistically significant
nr = not reported
na = not applicable

1. This appendix presents *Ladders to Literacy* group follow-up (first grade) data on measures that fall in the alphabetics domain. Findings that were used for rating purposes are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
6. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of O'Connor et al. (1996), corrections for clustering were needed. In case of Fuchs et al. (2001), no corrections were needed.
7. Student-level standard deviations were not available for this study. Thus, a student-level effect size and improvement index could not be computed, and are not presented in this table. Cluster-level standard deviations are presented in the author's findings columns.
8. The total number of students was 271 with approximately 40% low achievers, 30% average achievers, and 30 high achievers; however, the exact number in each group was not provided.

Appendix A5.1 *Ladders to Literacy* rating for the alphabetics domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of alphabetics, the WWC rated *Ladders to Literacy* as having potentially positive effects. It did not meet the criteria for positive effects because there were no studies that met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *Ladders to Literacy* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. Three studies showed statistically significant positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. There were no studies showing negative effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. There were no studies that met WWC standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. There were no studies showing negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Ladders to Literacy* rating for the fluency domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of fluency, the WWC rated *Ladders to Literacy* as having potentially positive effects. It did not meet the criteria for positive effects because there were no studies that met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *Ladders to Literacy* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. One study showed substantively important positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. There were no studies showing negative effects and one study showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. There were no studies showing statistically significant positive effects, and no studies that met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. There were no studies showing negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.3 *Ladders to Literacy* rating for the comprehension domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of comprehension, the WWC rated *Ladders to Literacy* as having mixed effects. It did not meet the criteria for positive effects because no studies showed statistically significant positive effects. In addition, it did not meet the criteria for potentially positive effects because two studies showed indeterminate effects and only one study showed a substantively important positive effect. The remaining ratings (no discernible effects, potentially negative effects, and negative effects) were not considered as *Ladders to Literacy* was assigned the highest applicable rating.

Rating received

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. One study of *Ladders to Literacy* had a substantively important positive effect in this domain. No studies showed a statistically significant or substantively important negative effect.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Met. One study of *Ladders to Literacy* had a substantively important positive effect in this domain and two studies showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No study of *Ladders to Literacy* had a statistically significant positive effect in this domain.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies of *Ladders to Literacy* showed statistically significant or substantively important negative effects in this domain.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. One study of *Ladders to Literacy* had a substantively important positive effect in this domain.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Not met. No studies of *Ladders to Literacy* showed statistically significant or substantively important negative effects in this domain, and more studies showed indeterminate effects (two) than substantively important positive effects (one) in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence rating by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	4	over 14	760	Moderate to large
Fluency	1	over 1	66	Small
Comprehension	3	over 6	489	Moderate to large
General reading achievement	0	0	0	na

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain, and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."